MANIFOLDS OF DIFFERENTIABLE DENSITIES

NIGEL J. NEWTON[*]

Abstract. We develop a family of infinite-dimensional (non-parametric) manifolds of probability measures. The latter are defined on underlying Banach spaces, and have densities of class C_b^k with respect to appropriate reference measures. The case $k = \infty$, in which the manifolds are modelled on Fréchet spaces, is included. The manifolds admit the Fisher-Rao metric and, unusually for the non-parametric setting, Amari's α-covariant derivatives for all $\alpha \in \mathbb{R}$. By construction, they are C^∞-embedded submanifolds of particular manifolds of finite measures. The statistical manifolds are dually ($\alpha = \pm 1$) flat, and admit mixture and exponential representations as charts. Their curvatures with respect to the α-covariant derivatives are derived. The likelihood function associated with a finite sample is a continuous function on each of the manifolds, and the α-divergences are of class C^∞.

1. INTRODUCTION

Information geometry is the study of differential-geometric structures arising in the theory of statistical estimation, and has a history going back (at least) to the work of Rao [28]. It is finding increasing application in many fields including *asymptotic statistics*, *machine learning*, *signal processing* and *statistical mechanics* (see, for example, [24, 25] for some recent developments). The theory in finite dimensions (the parametric case) is well developed, and treated pedagogically in a number of texts [1, 4, 9, 13, 19]. A classical example is the finite-dimensional *exponential model*, in which linear combinations of a finite number of real-valued random variables (defined on an underlying probability space $(\mathbb{X}, \mathcal{X}, \mu)$) are exponentiated and normalised to generate probability density functions with respect to the reference measure μ. The topology induced on the set of probability measures, thus defined, is consistent with the important statistical divergences of estimation theory, and derivatives of the latter can be used to define geometric objects such as a Riemannian metric (the Fisher-Rao metric) and a family of covariant derivatives.

Central to any extension of these ideas to infinite dimensions, is the use of charts with respect to which statistical divergences are sufficiently smooth. The *Kullback-Leibler divergence* between two probability measures $P \ll Q$ is defined as follows:

$$\mathcal{D}_{\mathrm{KL}}(P \,|\, Q) := \mathbf{E}_Q(\mathrm{d}P/\mathrm{d}Q) \log(\mathrm{d}P/\mathrm{d}Q), \tag{1.1}$$

where $\mathbf{E}_Q$ represents expectation (integration) with respect to Q. As is clear from (1.1), the regularity of $\mathcal{D}_{\mathrm{KL}}$ is closely connected with that of the density, $\mathrm{d}P/\mathrm{d}Q$, and its log (considered as elements of dual spaces of real-valued functions on $\mathbb{X}$). In fact, much of information geometry concerns the interplay between these two representations of P, and the exponential map that connects them. The two associated affine structures form the basis of a Fenchel-Legendre transform underpinning the subject, and so manifolds that fully accommodate these structures are particularly useful.

In the series of papers [8, 12, 26, 27], Pistone and his co-workers developed an infinite-dimensional variant of the exponential model outlined above. Probability measures in the manifold are mutually absolutely continuous with respect to the reference measure μ, and the manifold is covered by the charts $s_Q(P) = \log \mathrm{d}P/\mathrm{d}Q - \mathbf{E}_Q \log \mathrm{d}P/\mathrm{d}Q$ for different "patch-centric" probability measures Q. These readily give $\log \mathrm{d}P/\mathrm{d}Q$ the desired regularity, but require exponential Orlicz model spaces in order to do the same for $\mathrm{d}P/\mathrm{d}Q$. The exponential Orlicz manifold, M_O, has a strong topology, under which $\mathcal{D}_{\mathrm{KL}}$ is of class C^∞. In [26], the authors define "mean parameters" on M_O. Like s_Q, these are defined locally at each $Q \in M_O$, by $\eta_Q(P) = \mathrm{d}P/\mathrm{d}Q - 1$. η_Q maps into the "pre-dual" of the exponential Orlicz space. However, despite being injective, η_Q is not homeomorphic and so cannot be used as a chart. By contrast, the manifolds developed in [21, 23] use the "balanced" global chart $\phi(P) = \mathrm{d}P/\mathrm{d}\mu - 1 + \log \mathrm{d}P/\mathrm{d}\mu - \mathbf{E}_\mu \log \mathrm{d}P/\mathrm{d}\mu$, thereby enabling the use of model spaces with weaker topologies. (In order for $\mathcal{D}_{\mathrm{KL}}$ to be of class C^k, it suffices to use the Lebesgue model space $L^p(\mu)$ with $p = k+1$.) The balanced L^p manifold, M_B, admits *mixture* and *exponential* representations $m, e : M_B \to L^p(\mu)$, defined by $m(P) = \mathrm{d}P/\mathrm{d}\mu - 1$ and $e(P) = \log \mathrm{d}P/\mathrm{d}\mu - \mathbf{E}_\mu \log \mathrm{d}P/\mathrm{d}\mu$. Like η_Q on M_O, these are injective but not homeomorphic, and so cannot be used as charts. The Hilbert case, in which $p = 2$, is developed in detail in [21].

The exponential Orlicz and balanced L^p manifolds (for $p \geq 2$) support the infinite-dimensional variant of the Fisher-Rao metric, and (for $p \geq 3$) the infinite-dimensional variant of the *Amari-Chentsov tensor*. The latter can be used to define α-derivatives on particular *statistical bundles* [12]. However, with the exception of the case $\alpha = 1$ on the exponential Orlicz manifold, these bundles differ from the tangent bundle, and so the α-derivatives do not constitute covariant derivatives in the usual sense. This fact is intimately connected with the non-homeomorphic nature of η_Q on M_O, and m and e on M_B.

In [2], the authors define a very general notion of *statistical model*. This is a manifold equipped with a metric and symmetric 3-tensor, together with an embedding into a space of finite measures, such that these become the Fisher-Rao metric and Amari-Chentsov tensor. They extend a result of Chentsov (on the uniqueness of these tensors as invariants under sufficient statistics) to this much wider class of statistical models. The exponential Orlicz and balanced L^p manifolds (for $p \geq 3$) all fit within this framework.

The topologies of these manifolds (like those of all manifolds of "pure" information geometry) have no direct connection with any topology that the underlying space $(\mathbb{X}, \mathcal{X}, \mu)$ may possess. They concern statistical inference in its barest form – statistical divergences measure dependency between random variables without recourse to structures in their range spaces any richer than a σ-algebra of events. Nevertheless, metrics, topologies and linear structures on $\mathbb{X}$ play important roles in many applications. In *maximum likelihood estimation*, for example, it is desirable for the likelihood function associated with a finite sample to be continuous. It is, therefore, of interest to develop statistical manifolds that embrace both topologies. This is a central aim here; we incorporate the topology of $\mathbb{X}$ by using appropriate model space norms. A different approach is pursued in [11]. The exponential manifolds developed there admit, by construction, continuous evaluation maps (such as the likelihood function) since they are based on reproducing kernel Hilbert space methods. However, like the exponential Orlicz and balanced L^p manifolds, they do not fully accommodate the affine structure associated with the density.

In developing this work, the author was motivated by problems in Bayesian estimation, in which posterior distributions must be computed from priors and partial observations. Suppose, for example, that $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ are random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and taking values in metric spaces $\mathbb{X}$ and $\mathbb{Y}$, respectively. Let $\mathcal{X}$ be the σ-algebra of Borel subsets of $\mathbb{X}$, and let $\mathcal{P}(\mathcal{X})$ be the set of probability measures on $\mathcal{X}$. Under mild conditions we can construct a *regular conditional probability distribution* for X given Y, $\Pi : \mathbb{Y} \to \mathcal{P}(\mathcal{X})$ (see, for example, [15]). This has the key properties that, for each $B \in \mathcal{X}$, $\Pi(\cdot)(B) : \mathbb{Y} \to [0,1]$ is measurable and $P(X \in B|Y) = \Pi(Y)(B)$. In many instances of this problem, the image $\Pi(\mathbb{Y})$ is contained in an infinite-dimensional statistical manifold, and particular statistical divergences defined on the manifold

can be interpreted as "multi-objective" measures of error in approximations to Π. For example, Pearson's χ^2-divergence can be interpreted as a multi-objective measure of the normalised mean-squared errors in estimates of real-valued variates, $f(X)$ [23].

$$\mathcal{D}_{\chi^2}(P \,|\, Q) := \frac{1}{2}\|\mathrm{d}P/\mathrm{d}Q - 1\|_{L_0^2(Q)}^2 = \frac{1}{2}\sup_{f \in F}(\mathbf{E}_Q(\mathrm{d}P/\mathrm{d}Q - 1)f)^2 = \frac{1}{2}\sup_{f \in L^2(Q)}\frac{(\mathbf{E}_P f - \mathbf{E}_Q f)^2}{\mathbf{E}_Q(f - \mathbf{E}_Q f)^2}, \qquad (1.2)$$

where $L_0^2(Q)$ is the set of Q-square integrable, real-valued functions having mean zero, and F is the subset of those functions having unit variance. If $\hat{\Pi} : \mathbb{Y} \to \mathcal{P}(\mathcal{X})$ is used as an approximation to Π, and $\mathbf{E}_{\hat{\Pi}(Y)}f$ as an approximation of the variate $f(X)$, then the mean-squared error admits the orthogonal decomposition:

$$\mathbb{E}\left(f(X) - \mathbf{E}_{\hat{\Pi}(Y)}f\right)^2 = \mathbb{E}\mathbf{E}_{\Pi(Y)}\left(f - \mathbf{E}_{\Pi(Y)}f\right)^2 + \mathbb{E}\left(\mathbf{E}_{\hat{\Pi}(Y)}f - \mathbf{E}_{\Pi(Y)}f\right)^2. \qquad (1.3)$$

The first term on the right-hand side here is the *statistical error* arising from the limitations of the observation Y, whereas the second term is the approximation error arising from the use of $\hat{\Pi}$ instead of Π. Since there is no point in approximating $\mathbf{E}_{\Pi(Y)}f$ with great accuracy if it is itself a poor estimate of $f(X)$, it is appropriate to consider the magnitude of the second term *relative* to that of the first. As the final term in (1.2) shows, the χ^2-divergence, $\mathcal{D}_{\chi^2}(\hat{\Pi}(Y) \,|\, \Pi(Y))$, selects the worst of these relative errors among the square-integrable variates $f(X)$. The divergences $\mathcal{D}_{KL}$ and $\mathcal{D}_{\chi^2}$ are both members of the one-parameter family of α-*divergences*, $(\mathcal{D}^\alpha, \alpha \in \mathbb{R})$ [1]. (In fact $\mathcal{D}_{KL} = \mathcal{D}^{-1}$ and $\mathcal{D}_{\chi^2} = \mathcal{D}^{-3}$.)

Bayesian problems, in which a Markov process $(X_t, 0 \le t < \infty)$ has to be estimated at each time t, on the basis of the history of an observation process $(Y_s, 0 \le s \le t)$, are known as problems in *nonlinear filtering*. Regular conditional distributions for X are then time dependent, and can often be represented as solutions of stochastic partial differential equations [15]. Suppose, for example, that X is an $\mathbb{R}^d$-valued diffusion process with drift coefficient $b : \mathbb{R}^d \to \mathbb{R}^d$ and positive semi-definite diffusion coefficient $a : \mathbb{R}^d \to \mathbb{R}^{d \times d}$. Suppose, further, that Y is a real-valued partial observation process of the type

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + W_t, \quad 0 \le t < \infty, \qquad (1.4)$$

where $h : \mathbb{R}^d \to \mathbb{R}$ is a measurable function and W is a Brownian motion, independent of X. Under suitable regularity constraints on a, b and h, the $(Y_s, 0 \le s \le t)$-conditional distribution of X_t has a density π_t satisfying the *Kushner-Stratonovich equation* [15]:

$$\pi_t = \pi_0 + \int_0^t \mathcal{A}\pi_s\,\mathrm{d}s + \int_0^t \pi_s(h - \bar{h}_s)(\mathrm{d}Y_s - \bar{h}_s\,\mathrm{d}s), \qquad (1.5)$$

where $\mathcal{A}$ is the Kolmogorov forward (Fokker-Planck) operator for X, and $\bar{h}_s$ is the conditional mean of h:

$$\mathcal{A}f = \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2(a_{ij}f)}{\partial x_i \partial x_j} - \sum_{i=1}^d \frac{\partial(b_i f)}{\partial x_i} \quad \text{and} \quad \bar{h}_s = \int h(x)\pi_s(x)\,\mathrm{d}x. \qquad (1.6)$$

If posterior distributions of a nonlinear filter stay on a statistical manifold, then it is possible to use the methods of information geometry to study its information-theoretic properties, and to develop approximations based on finite-dimensional submanifolds. These ideas are developed in [6] in the context of the exponential Orlicz manifold, and in [22] in the context of the balanced Hilbert manifold. However these manifolds are not suited to the quest for infinite dimensional evolution equations since they are constructed without reference to the topology of $\mathbb{R}^d$, which is clearly needed in any vector field representation of the first integral in (1.5).

To overcome this problem, we need a statistical manifold with a model space whose members satisfy suitable differentiability constraints. In [17], the authors define an infinite-dimensional statistical manifold modelled on a weighted Orlicz-Sobolev space, and use it to study the Boltzmann equation. This manifold can also be used to study weak solutions of (1.5).

The two integral terms on the right-hand side of (1.5) are *mixture affine* (respectively *exponential affine*), in the sense that the integrands are affine maps in the η_Q/m (respectively s_Q/e) representations. In the quest for evolution equations, it is therefore advantageous to use a manifold that admits both these representations as charts. Such a manifold is constructed here; it comprises probability measures whose log-densities with respect to a reference measure are of class C_b^k. In particular, their densities have strictly positive infima, which is a significant restriction if, for example, $\mathbb{X} = \mathbb{R}^d$. This constraint can be thought of as an infinite-dimensional equivalent of the positivity constraint placed on the probabilities of all atoms in the finite sample space setting (see Sect. 2.5 in [1]). As in that setting, the removal of this constraint would add a boundary to the manifold, on which certain divergences are singular. The manifolds constructed are suited to nonlinear filtering problems in which the process X is constrained to lie in a bounded subset of $\mathbb{R}^d$ by certain types of boundary condition, such as reflective (Neumann) or periodic conditions (see Rem. 2.2(iv)). If the boundaries are sufficiently far from the origin, then problems with boundaries may be just as good models for physical systems as those without. If, for example, the drift and diffusion coefficients b and a are bounded, and a is strictly positive definite, then posterior densities are known to have Gaussian tails. Such a density may be no more accurate a representation of reality 10 standard deviations from its mean than a density coming from a model that incorporates a reflective boundary at this point.

Aside from applications to nonlinear filtering, it is of fundamental interest to develop non-parametric statistical manifolds that admit the full geometry of Amari's α-covariant derivatives – something that is not achieved in the manifolds described above.

The paper is structured as follows. Sections 2 and 3 construct M, a smooth manifold of *finite* measures on an open subset of a *Banach space* $\mathbb{X}$, whose densities with respect to a reference measure are of class C_b^k. M is covered by each chart in a one-parameter family $(\phi_\alpha, \alpha \in \mathbb{R})$. The charts ϕ_α and $\phi_{-\alpha}$ map to open subsets of the affine spaces of a Fenchel-Legendre transform involving the α-divergences $\mathcal{D}_\alpha$ and $\mathcal{D}_{-\alpha}$. As such, they define a metric and dual notions of parallel transport on the tangent bundle for each $\alpha \in \mathbb{R}$. Since all these charts are global, M is flat with respect to the associated covariant derivatives, $(\nabla^\alpha, \alpha \in \mathbb{R})$. Section 4 considers the subset of *probability* measures, N. This is a C^∞-embedded submanifold of M, from which it inherits its important properties. In particular, the projection of the metric and covariant derivatives of M onto N yields the Fisher-Rao metric, and the α-covariant derivatives on N. In contrast with the manifolds of [8, 12, 17, 21, 23, 26, 27], the latter are all defined on the tangent bundle of N. Of course, this extra regularity is gained at the cost of inclusiveness. N is (dually) flat in the $\alpha = \pm 1$-covariant derivatives, and admits affine mixture and exponential charts. Finally, Section 5 uses the method of projective limits to extend these results to manifolds of *smooth* densities. The idea of embedding non-parametric statistical manifolds in manifolds of finite measures is not new (see, for example, [2] and [23]). However, the fact that the ambient manifold here admits a multiplication operator ((2.2) below) allows the global α-geometry of the statistical manifold to be obtained in an extrinsic manner.

In recent related work [5, 7], the authors construct a manifold of smooth densities on an underlying finite-dimensional manifold by considering such densities to be smooth sections of the associated *volume bundle*. (This is a vector bundle of dimension 1 that endows the underlying manifold with an intrinsic notion of volume.) They consider a property of invariance of Riemannian metrics under the diffeomorphism group of the underlying manifold, and construct the class of all metrics with this property. When restricted to the submanifold of *probability measures*, these all coincide (modulo scaling) with the Fisher-Rao metric. In [7], they develop the Levi-Civita covariant derivative and carry out a number of extensions and completions of the manifold in order to study its global geometry.

2. The exponential map

Let B be an open subset of a Banach space $\mathbb{X}$, and let $\mathcal{X}_B := \{A \subset B : A \in \mathcal{X}\}$, where $\mathcal{X}$ is the Borel σ-algebra on $\mathbb{X}$. Let μ be a probability measure on $(B, \mathcal{X}_B)$ with the following property: for any non-empty open $A \in \mathcal{X}_B$, $\mu(A) > 0$. (For example, $\mathbb{X} = \mathbb{R}^d$, B is a bounded open rectangle, and μ is normalised Lebesgue measure.) Let $\mathbb{G} := C_b^k(B; \mathbb{R})$ be the space of continuous and bounded functions $a : B \to \mathbb{R}$, that have continuous and bounded (Fréchet) derivatives of all orders up to some $k \in \mathbb{N}_0$. $\mathbb{G}$ is a Banach space over $\mathbb{R}$ when endowed with the norm:

$$\|a\|_{\mathbb{G}} = \sup_{x \in B} |a(x)| + \sum_{i=1}^{k} \sup_{x \in B} \|a_x^{(i)}\|_{L(\mathbb{X}^i; \mathbb{R})}, \tag{2.1}$$

where $a^{(i)} : B \to L(\mathbb{X}^i; \mathbb{R})$ is the i'th derivative of a, and $L(\mathbb{X}^i; \mathbb{R})$ is the Banach space of continuous multilinear functions from $\mathbb{X}^i$ to $\mathbb{R}$, endowed with the operator norm. The (continuous bilinear) *multiplication operator* $\pi : \mathbb{G} \times \mathbb{G} \to \mathbb{G}$, and the (continuous linear) *expectation functional* $\mathbf{E}_\mu : \mathbb{G} \to \mathbb{R}$, are as follows

$$(a \cdot b)(x) = \pi(a, b)(x) = a(x)b(x) \quad \text{and} \quad \mathbf{E}_\mu a = \int_B a(x)\mu(\mathrm{d}x). \tag{2.2}$$

Equipped with π, $\mathbb{G}$ becomes a commutative Banach algebra with identity $\mathbf{e} \equiv 1$. In the special case that $k = 0$, it is a commutative C^*-algebra with involution the identity map.

Proposition 2.1. *The Nemytskii (superposition) operator, $\exp_{\mathbb{G}} : \mathbb{G} \to \mathbb{G}^+$, defined by $\exp_{\mathbb{G}}(a)(x) = \exp_{\mathbb{R}}(a(x))$, is diffeomorphic onto its image $\mathbb{G}^+ := \{a \in \mathbb{G} : \inf_{x \in B} a(x) > 0\}$, and has first derivative*

$$\exp_{\mathbb{G},a}^{(1)} u = \exp_{\mathbb{G}}(a) \cdot u. \tag{2.3}$$

Proof. Let $F : \mathbb{G} \times \mathbb{G} \to \mathbb{G}$ be defined by $F(a, b) = \exp_{\mathbb{G}}(b) - \exp_{\mathbb{G}}(a) - \exp_{\mathbb{G}}(a) \cdot (b - a)$. In order to prove (2.3) it suffices to show that, for any $a \in \mathbb{G}$, there exists a $K_a < \infty$ such that

$$\|F(a, b)\|_{\mathbb{G}} \le K_a \|b - a\|_{\mathbb{G}}^2 \quad \text{for all } b \in B(a, 1), \tag{2.4}$$

where $B(a, 1)$ is the open unit ball centred at a. That this is so when $k = 0$ follows from an application of Taylor's theorem to $\exp_{\mathbb{R}}$. Suppose that $k \ge 1$. Fixing $a \ne b \in \mathbb{G}$, and differentiating $F(a, b)$ with respect to x, we obtain

$$F(a, b)_x^{(1)} y = F(a, b)(x)b_x^{(1)} y + H(a, b, x)y, \tag{2.5}$$

where $H : \mathbb{G} \times \mathbb{G} \times B \to L(\mathbb{X}; \mathbb{R})$ is defined by

$$H(a, b, x)y = (\exp_{\mathbb{G}}(a) \cdot (b - a))(x)(b_x^{(1)} - a_x^{(1)})y.$$

An induction argument, starting from (2.5), shows that, for any $1 \le i \le k$,

$$F(a, b)_x^{(i)} \mathbf{y}_1^i = \sum_{\rho \in S_i} \sum_{j=1}^{i} \gamma_{i,\rho,j} F(a, b)_x^{(i-j)} \mathbf{y}_{\rho_{j+1}}^{\rho_i} b_x^{(j)} \mathbf{y}_{\rho_1}^{\rho_j} + H(a, b, \cdot)_x^{(i-1)} \mathbf{y}_1^i, \tag{2.6}$$

where $\mathbf{y}_m^n := (y_m, \ldots, y_n)$, S_i is the set of all permutations of the integers 1 to i, and the real constants $\gamma_{i,\rho,j}$ are defined by the following recursion: $\gamma_{i,\rho,0} = \gamma_{i,\rho,i+1} = 0$, $\gamma_{1,\mathbf{e},1} = 1$, and for any $\sigma \in S_{i+1}$ and any $1 \le j \le i+1$,

$$\gamma_{i+1,\sigma,j} = \begin{cases} \gamma_{i,\rho,j} & \text{if } \sigma = (\rho, i+1) \text{ for some } \rho \in S_i \\ \gamma_{i,\rho,j-1} & \text{if } \sigma = (\rho_1, \ldots, \rho_{j-1}, i+1, \rho_j, \ldots, \rho_i) \text{ for some } \rho \in S_i \\ 0 & \text{otherwise.} \end{cases}$$

For any $a \in \mathbb{G}$, there exists a $K_a < \infty$ such that

$$\sup_{x \in B} \|H(a, b, \cdot)_x^{(i-1)}\|_{L(\mathbb{X}^i; \mathbb{R})} \le K_a \|b - a\|_{\mathbb{G}}^2 \quad \text{for all } b \in \mathbb{G}. \tag{2.7}$$

An induction argument on i thus establishes (2.4), and hence (2.3). A further induction argument readily shows that $\exp_{\mathbb{G}} \in C^\infty(\mathbb{G}; \mathbb{G}^+)$.

For any $a \in \mathbb{G}$, the linear map $\exp_{\mathbb{G},a}^{(1)} : \mathbb{G} \to \mathbb{G}$ of (2.3) is clearly a toplinear isomorphism, and so the statement of the proposition follows from the inverse mapping theorem. □

Remark 2.2.

(i) The crucial feature of this setup is that $\exp_{\mathbb{G}}(\mathbb{G})$ is an open subset of $\mathbb{G}$, which is essential if both mixture and exponential representations are to be charts. This property, which is connected with the existence of the multiplication operator of (2.2), does not hold if $\mathbb{G}$ is replaced by the exponential Orlicz spaces of [17, 27], or the Lebesgue $L^p(\mu)$ spaces of [21, 23] (see examples 2.1 and 2.2 in [21]).

(ii) The model space $\mathbb{G}$ is based on an *open* subset of $\mathbb{X}$ for reasons of inclusiveness. The closure of B, $\bar{B}$, does not need to be compact. If, however, $\bar{B}$ *is* compact, then the space $\{a \in \mathbb{G} : a = b|_B \text{ for some } b \in C^k(\bar{B}; R)\}$ is a proper, closed subspace of $\mathbb{G}$, and so defines smoothly embedded submanifolds of those constructed in Sections 3 and 4.

(iii) The measure μ on $(B, \mathcal{X}_B)$ must be *finite* in order for the integral functional of (2.2) to be well defined. Since the total mass of μ does not affect the results that follow, it is natural to assume that it is 1. μ is then, itself, a member of the statistical manifold of Section 4.

(iv) The results that follow hold true in other scenarios. For example, that in which $\mathbb{X} = \mathbb{R}^d$, $B = (-\pi, \pi)^d$, $\mu = (2\pi)^{-d}\text{Leb}$ and $\mathbb{G}$ is the subspace of $C_b^k(B; \mathbb{R})$ whose members satisfy a suitable periodic boundary condition. The manifolds constructed then comprise measures defined on the d-dimensional torus.

(v) $\mathbb{G}$ can also be replaced by $L^\infty(B; \mathbb{R})$, but no account is then taken of the topology of $\mathbb{X}$. *cf.* [16], in which the authors follow the approach of [27] to construct a *Tsallis q-exponential* statistical manifold modelled on a Banach space of essentially bounded functions.

3. The manifold of finite measures

Let M be the set of finite measures on $(B, \mathcal{X}_B)$ that are mutually absolutely continuous with respect to μ, and have densities of the form

$$\mathrm{d}P/\mathrm{d}\mu = \exp_{\mathbb{G}}(a), \quad \text{for some } a \in \mathbb{G}. \tag{3.1}$$

M is covered by the single chart $\phi_1 : M \to \mathbb{G}$, defined by $\phi_1(P) = \log_{\mathbb{G}^+}(dP/d\mu) := \exp_{\mathbb{G}}^{-1}(dP/d\mu)$. A tangent vector at $P \in M$ is a signed measure on $(B, \mathcal{X}_B)$ that is absolutely continuous with respect to μ, and has a density with respect to P of the form

$$\mathrm{d}U/\mathrm{d}P = u, \quad \text{for some } u \in \mathbb{G}. \tag{3.2}$$

The tangent space at P, $T_P M$, is the linear space of all such measures, and the tangent bundle is the disjoint union $TM := \bigcup_{P \in M}(P, T_P M)$. By construction, TM is globally trivialised by the bijection $\Phi_1 : TM \to \mathbb{G} \times \mathbb{G}$

defined by

$$\Phi_1(P, U) = (\log_{\mathbb{G}^+} \mathrm{d}P/\mathrm{d}\mu, \, \mathrm{d}U/\mathrm{d}P). \tag{3.3}$$

The derivative of a differentiable, Banach-space-valued map $f : M \to \mathbb{Y}$ (at P and in the direction $U \in T_P M$) is defined in the obvious way:

$$Uf := (f \circ \phi_1^{-1})_a^{(1)} u, \quad \text{where } (a, u) = \Phi_1(P, U). \tag{3.4}$$

For any $\alpha \in \mathbb{R} \setminus \{1\}$, let $\phi_\alpha : M \to \mathbb{G}$ be defined as follows:

$$\phi_\alpha(P) = \tfrac{2}{1-\alpha} \left(\exp_{\mathbb{G}} \left(\tfrac{1-\alpha}{2} \phi_1(P) \right) - 1 \right). \tag{3.5}$$

Proposition 2.1 shows that the map $\phi_\alpha \circ \phi_1^{-1}$ is diffeomorphic onto its image, and so $(\phi_\alpha, \, \alpha \in \mathbb{R})$ is a smooth atlas, each chart of which covers M. For any $\alpha, \beta \in \mathbb{R}$, the derivative of the transition map $\phi_\alpha \circ \phi_\beta^{-1}$ is,

$$(\phi_\alpha \circ \phi_\beta^{-1})_a^{(1)} u = \exp_{\mathbb{G}} \left(\tfrac{\beta - \alpha}{2} a_1 \right) \cdot u, \tag{3.6}$$

where $a_1 = \phi_1 \circ \phi_\beta^{-1}(a)$. For each $\alpha \in \mathbb{R}$ the chart $\Phi_\alpha : TM \to \mathbb{G} \times \mathbb{G}$, defined by

$$\Phi_\alpha(P, U) = (\phi_\alpha(P), U\phi_\alpha), \tag{3.7}$$

induces a distinct global trivialisation of TM. In particular, $\Phi_{-1}(P, U) = (\mathrm{d}P/\mathrm{d}\mu - 1, \mathrm{d}U/\mathrm{d}\mu)$.

Remark 3.1.

(i) The maps ϕ_α are derived from Amari's *α-embedding maps* (see Sect. 2.6 in [1]). The offset -1 is included in (3.5) so that $\phi_\alpha(\mu) = 0$. This also ensures that $\phi_\alpha \circ \phi_{-1}^{-1} \circ (\mathrm{id}_{\mathbb{G}^+} -1) : \mathbb{G}^+ \to \mathbb{G}$ is Tsallis' *q-deformed logarithm* with $q = (1 + \alpha)/2$ (see, for example, chapter 7 of [20]). It is easily established that, for any fixed $P \in M$, the map $\mathbb{R} \ni \alpha \mapsto \phi_\alpha(P) \in \mathbb{G}$ is of class C^∞.

(ii) We introduce multiple charts in order to define different notions of parallel transport on TM. Maps simlar to $(\phi_\alpha, \alpha \in [-1, 1])$ are defined and studied on the exponential Orlicz manifold in [12], and on the balanced $L^p(\mu)$ manifolds in [23]. However, since they are not diffeomorphic in those contexts, it is not possible to use them to define parallel transport on the associated tangent bundles.

(iii) The charts ϕ_{-1} and ϕ_1 are particularly important. ϕ_{-1} reflects the inherent linear structure of a set of measures. On the other hand, ϕ_1 is surjective, and so trivially introduces a Lie group structure on M. For $P, Q \in M$, the product PQ and power P^λ (for any $\lambda \in \mathbb{R}$) are defined as follows:

$$\frac{dPQ}{\mathrm{d}\mu} = \frac{\mathrm{d}QP}{\mathrm{d}\mu} = \left(\frac{\mathrm{d}P}{\mathrm{d}\mu} \cdot \frac{\mathrm{d}Q}{\mathrm{d}\mu} \right) \quad \text{and} \quad \frac{\mathrm{d}P^\lambda}{\mathrm{d}\mu} = \left(\frac{\mathrm{d}P}{\mathrm{d}\mu} \right)^\lambda, \tag{3.8}$$

and the identity is μ. The *power of a measure* for $\lambda \in (0, 1]$ is used in [3] to characterise the Fisher-Rao metric and Amari-Chentsov tensor on parametric statistical manifolds admitting singular measures. Since the log densities, here, are bounded, P^λ is defined for all real λ.

Let ΓTM be the space of smooth sections of TM (*i.e.* smooth vector fields). Each chart Φ_α induces a notion of parallel transport on TM; tangent vectors in different fibres of TM, $U \in T_P M$ and $\tilde{U} \in T_Q M$, are α-parallel transports of each other if $U\phi_\alpha = \tilde{U}\phi_\alpha$. The associated covariant derivative, $\nabla^\alpha : \Gamma TM \times \Gamma TM \to \Gamma TM$, is

that for which ϕ_α is an affine chart:

$$\nabla_{\mathbf{U}}^{\alpha}\mathbf{V}\phi_\alpha = \mathbf{U}\mathbf{V}\phi_\alpha. \tag{3.9}$$

As is the case for all covariant derivatives defined from global affine charts, ∇^α is torsion free, and M is ∇^α-flat (or simply α-flat) for all $\alpha \in \mathbb{R}$. In fact, for any $\mathbf{U}, \mathbf{V}, \mathbf{W} \in \Gamma TM$ and any $\alpha \in \mathbb{R}$,

$$\left(\nabla_{\mathbf{U}}^{\alpha}\mathbf{V} - \nabla_{\mathbf{V}}^{\alpha}\mathbf{U} - [\mathbf{U}, \mathbf{V}]\right)\phi_\alpha = \mathbf{U}\mathbf{V}\phi_\alpha - \mathbf{V}\mathbf{U}\phi_\alpha - [\mathbf{U}, \mathbf{V}]\phi_\alpha = 0,$$

$$\left(\nabla_{\mathbf{U}}^{\alpha}\nabla_{\mathbf{V}}^{\alpha}\mathbf{W} - \nabla_{\mathbf{V}}^{\alpha}\nabla_{\mathbf{U}}^{\alpha}\mathbf{W} - \nabla_{[\mathbf{U},\mathbf{V}]}^{\alpha}\mathbf{W}\right)\phi_\alpha = \mathbf{U}\mathbf{V}\mathbf{W}\phi_\alpha - \mathbf{V}\mathbf{U}\mathbf{W}\phi_\alpha - [\mathbf{U}, \mathbf{V}]\mathbf{W}\phi_\alpha = 0, \tag{3.10}$$

where $[\cdot, \cdot]$ is the Lie bracket. α-geodesics are curves of M whose ϕ_α-representations are straight lines in $\mathbb{G}$.

We define a weak Riemannian metric on M *via* the inclusion $\mathbb{G} \subset L^2(P)$: for any $U, V \in T_P M$,

$$\langle U, V \rangle_P := \langle U\phi_1, V\phi_1 \rangle_{L^2(P)} = \langle U\phi_\alpha, V\phi_{-\alpha} \rangle_{L^2(\mu)} \quad \text{for all } \alpha \in \mathbb{R}, \tag{3.11}$$

where we have used (3.6) in the second step. This is positive definite since $P(A) > 0$ for any non-empty open set $A \in \mathcal{X}_B$. (It is not a *strong* Riemannian metric since there are members of the cotangent space that do not admit the representation $\langle U, \cdot \rangle_P : T_P M \to \mathbb{R}$ for some $U \in T_P M$.)

As is clear from (3.11), if $\tilde{U}, \tilde{V} \in T_Q M$ are obtained by parallel transport of $U, V \in T_P M$, one according Φ_α and the other according to $\Phi_{-\alpha}$, then $\langle \tilde{U}, \tilde{V} \rangle_Q = \langle U, V \rangle_P$. In this sense ∇^α and $\nabla^{-\alpha}$ are *dual* with respect to the metric. This can be expressed in differential form as follows: for any $\mathbf{U}, \mathbf{V}, \mathbf{W} \in \Gamma TM$,

$$\begin{aligned}
\mathbf{U}\langle \mathbf{V}, \mathbf{W} \rangle &= \mathbf{U}\langle \mathbf{V}\phi_\alpha, \mathbf{W}\phi_{-\alpha} \rangle_{L^2(\mu)} \\
&= \langle \mathbf{U}\mathbf{V}\phi_\alpha, \mathbf{W}\phi_{-\alpha} \rangle_{L^2(\mu)} + \langle \mathbf{V}\phi_\alpha, \mathbf{U}\mathbf{W}\phi_{-\alpha} \rangle_{L^2(\mu)} \\
&= \langle \nabla_{\mathbf{U}}^{\alpha}\mathbf{V}, \mathbf{W} \rangle + \langle \mathbf{V}, \nabla_{\mathbf{U}}^{-\alpha}\mathbf{W} \rangle.
\end{aligned} \tag{3.12}$$

(*cf.* the finite-dimensional case [1].) Being self-dual and torsion free, ∇^0 is the Levi-Civita covariant derivative associated with the metric. The linear relation between the α-covariant derivatives is also retained:

$$\nabla^\alpha = \tfrac{1-\alpha}{2}\nabla^{-1} + \tfrac{1+\alpha}{2}\nabla^1. \tag{3.13}$$

This follows from (3.6), which shows that

$$\nabla_{\mathbf{U}}^{\pm 1}\mathbf{V}\phi_\alpha = (\phi_\alpha \circ \phi_{\pm 1}^{-1})_{\phi_{\pm 1}}^{(1)}\nabla_{\mathbf{U}}^{\pm 1}\mathbf{V}\phi_{\pm 1} = (\phi_\alpha \circ \phi_{\pm 1}^{-1})_{\phi_{\pm 1}}^{(1)}\mathbf{U}\left[(\phi_{\pm 1} \circ \phi_\alpha^{-1})_{\phi_\alpha}^{(1)}\mathbf{V}\phi_\alpha\right] = \mathbf{U}\mathbf{V}\phi_\alpha + \tfrac{\alpha \mp 1}{2}\mathbf{U}\phi_1 \cdot \mathbf{V}\phi_\alpha.$$

3.1. The α-divergences

These are defined on M as follows. (*cf.* Sect. 3.6 in [1].)

$$\mathcal{D}_{-1}(P \,|\, Q) = \mathcal{D}_1(Q \,|\, P) = \mathbf{E}_\mu(\phi_{-1}(Q) - \phi_{-1}(P)) + \langle \phi_{-1}(P) + 1, \phi_1(P) - \phi_1(Q) \rangle_{L^2(\mu)}, \tag{3.14}$$

and, for $\alpha \neq \pm 1$,

$$\mathcal{D}_\alpha(P \,|\, Q) = \tfrac{2}{1+\alpha}\mathbf{E}_\mu(\phi_{-1}(P) - \phi_\alpha(P)) + \tfrac{2}{1-\alpha}\mathbf{E}_\mu(\phi_{-1}(Q) - \phi_{-\alpha}(Q)) - \langle \phi_\alpha(P), \phi_{-\alpha}(Q) \rangle_{L^2(\mu)}. \tag{3.15}$$

It follows from Proposition 2.1 that $\mathcal{D}_\alpha \in C^\infty(M \times M; \mathbb{R})$. The following proposition summarises some other properties.

Proposition 3.2. *For any $\alpha \in \mathbb{R}$:*

(i) $\mathcal{D}_{-\alpha}(P \,|\, Q) = \mathcal{D}_\alpha(Q \,|\, P) \geq 0$, *with equality if and only if $P = Q$;*

(ii) *the following* generalised cosine rule *applies*

$$\mathcal{D}_\alpha(P \,|\, R) = \mathcal{D}_\alpha(P \,|\, Q) + \mathcal{D}_\alpha(Q \,|\, R) - \langle \phi_\alpha(P) - \phi_\alpha(Q), \phi_{-\alpha}(R) - \phi_{-\alpha}(Q) \rangle_{L^2(\mu)}; \qquad (3.16)$$

(iii) *the set $\phi_\alpha(M)$ is convex;*

(iv) *for any $Q \in M$, the function $\mathcal{D}_\alpha(\phi_\alpha^{-1} |\, Q) : \phi_\alpha(M) \to \mathbb{R}$ admits the following derivatives*

$$\mathcal{D}_\alpha(\phi_\alpha^{-1} |\, Q)_a^{(1)} u = \langle \phi_{-\alpha} \circ \phi_\alpha^{-1}(a) - \phi_{-\alpha}(Q), u \rangle_{L^2(\mu)} \qquad (3.17)$$

$$\mathcal{D}_\alpha(\phi_\alpha^{-1} |\, Q)_a^{(2)}(u, v) = \mathbf{E}_\mu \exp_{\mathbb{G}}(\alpha \phi_1 \circ \phi_\alpha^{-1}(a)) \cdot u \cdot v; \qquad (3.18)$$

in particular $\mathcal{D}_\alpha(\phi_\alpha^{-1} |\, Q)$ is strictly convex;

(v) *for any $Q \in M$, $a \in \phi_{-\alpha}(M)$,*

$$\mathcal{D}_{-\alpha}(\phi_{-\alpha}^{-1}(a) \,|\, Q) = \max_{b \in \phi_\alpha(M)} \left\{ \langle a - \phi_{-\alpha}(Q), b - \phi_\alpha(Q) \rangle_{L^2(\mu)} - \mathcal{D}_\alpha(\phi_\alpha^{-1}(b) \,|\, Q) \right\}, \qquad (3.19)$$

and the unique maximiser is $\phi_\alpha \circ \phi_{-\alpha}^{-1}(a)$.

Proof. Parts (i), (ii) and (iv) can be proven by straightforward calculations. Part (iii) is trivial when $\alpha = 1$, since $\phi_1(M) = \mathbb{G}$. Suppose, then, that $\alpha \in \mathbb{R} \setminus \{1\}$. For any distinct $P_0, P_1 \in M$, and any $t \in (0, 1)$, let $a_t := (1 - t)\phi_\alpha(P_0) + t\phi_\alpha(P_1)$; then we can define

$$p_t = \left(1 + \tfrac{1-\alpha}{2} a_t \right)^{2/(1-\alpha)} = \left((1 - t)p_0^{(1-\alpha)/2} + t p_1^{(1-\alpha)/2} \right)^{2/(1-\alpha)},$$

where $p_i := \mathrm{d}P_i/\mathrm{d}\mu$, $i = 0, 1$. Since the infimum (over $x \in B$) of the term in brackets on the right-hand side here is strictly positive, $\log p_t$ is well defined and bounded. p_t is thus the density of a measure $P_t \in M$, and $\phi_\alpha(P_t) = a_t$, which completes the proof of part (iii).

Let $a \in \phi_{-\alpha}(M)$, and let $f : \phi_\alpha(M) \to \mathbb{R}$ be defined as follows:

$$\begin{aligned}
f(b) &= \langle a - \phi_{-\alpha}(Q), b - \phi_\alpha(Q) \rangle_{L^2(\mu)} - \mathcal{D}_\alpha(\phi_\alpha^{-1}(b) \,|\, Q) \\
&= \mathcal{D}_\alpha(Q \,|\, \phi_{-\alpha}^{-1}(a)) - \mathcal{D}_\alpha(\phi_\alpha^{-1}(b) \,|\, \phi_{-\alpha}^{-1}(a)) \\
&= \mathcal{D}_{-\alpha}(\phi_{-\alpha}^{-1}(a) \,|\, Q) - \mathcal{D}_{-\alpha}(\phi_{-\alpha}^{-1}(a) \,|\, \phi_\alpha^{-1}(b)),
\end{aligned}$$

where we have used (3.16) in the second step and part (i) in the third step. Part (v) now follows from part (i). $\qquad \square$

It follows from (3.11), (3.18) and (3.6) that, for any $P \in M$ and $U, V \in T_P M$,

$$\langle U, V \rangle_P = \mathcal{D}_\alpha(\phi_\alpha^{-1} |\, P)_{\phi_\alpha(P)}^{(2)}(U\phi_\alpha, V\phi_\alpha), \qquad (3.20)$$

confirming the Hessian nature of the metric. Furthermore, $\mathcal{D}_\alpha(\phi_\alpha^{-1} |\, \phi_\alpha^{-1})_{a,a}^{(2,1)}(u, v; \cdot) \equiv 0$ for all $(a, u), (a, v) \in \Phi_\alpha(TM)$. Since the metric is positive definite, the only tangent vector $X \in T_P M$, for which

$$\langle X, W \rangle_P = \mathcal{D}_\alpha(\phi_\alpha^{-1} |\, \phi_\alpha^{-1})_{a,a}^{(2,1)}(u, v; W\phi_\alpha) \quad \text{for all } W \in T_P M, \qquad (3.21)$$

is the zero vector, and this provides further justification for the definition of the covariant derivative in (3.9). (*cf.* the finite-dimensional case in Sect. 3.2 of [1].)

4. The manifold of probability measures

Let $\mathbb{G}_0 := \{a \in \mathbb{G} : \mathbf{E}_\mu a = 0\}$, let $N := \{P \in M : P(B) = 1\}$, and let $\phi_m : N \to \mathbb{G}_0$ be the restriction of ϕ_{-1} to N. N is a statistical manifold modelled on $\mathbb{G}_0$, with global *mixture chart* ϕ_m. It is trivially a C^∞-embedded submanifold of M. A tangent vector at $P \in N$ is a signed measure in $T_P M$ that has total mass zero. The tangent bundle, TN, is trivialised by the global chart $\Phi_m : TN \to \mathbb{G}_0 \times \mathbb{G}_0$, defined to be the restriction of Φ_{-1} to TN. We can also define an *exponential chart*, $\Phi_e : TN \to \mathbb{G}_0 \times \mathbb{G}_0$, as follows:

$$\Phi_e(P, U) = (\phi_e(P), U\phi_e) = (\phi_1(P) - \mathbf{E}_\mu \phi_1(P), \, \mathrm{d}U/\mathrm{d}P - \mathbf{E}_\mu \mathrm{d}U/\mathrm{d}P). \tag{4.1}$$

M and N are connected by the *normalisation map* $\nu : M \to N$ ($\nu(P) := P/P(B)$) and the *inclusion map*, $\imath : N \to M$. The associated tangent maps, $T\nu$ and $T\imath$, have particularly simple representations in terms of the charts Φ_1 and Φ_e:

$$\begin{aligned}
\Phi_e \circ T\nu \circ \Phi_1^{-1}(a, u) &= (a - \mathbf{E}_\mu a, u - \mathbf{E}_\mu u) \\
\Phi_1 \circ T\imath \circ \Phi_e^{-1}(b, v) &= (b - \log \mathbf{E}_\mu \exp_{\mathbb{G}}(b), v - \mathbf{E}_P v),
\end{aligned} \tag{4.2}$$

where $\mathbf{E}_P$ is expectation with respect to $P = \phi_e^{-1}(b)$. So

$$\Phi_1 \circ T\imath \circ T\nu \circ \Phi_1^{-1}(a, u) = (a - \log \mathbf{E}_\mu \exp_{\mathbb{G}}(a), u - \mathbf{E}_P u), \tag{4.3}$$

where $P = \nu \circ \phi_1^{-1}(a)$. A tangent vector $V \in T_P M$ at $P \in N$ is in $T_P N$ if and only if $\mathbf{E}_P V\phi_1 = \mathbf{E}_P \mathrm{d}U/\mathrm{d}P = 0$. So, for any $P \in N$, $U \in T_P M$ and $V \in T_P N$,

$$\langle U, V \rangle_P = \langle U\phi_1, V\phi_1 \rangle_{L^2(P)} = \langle U\phi_1 - \mathbf{E}_P U\phi_1, V\phi_1 \rangle_{L^2(P)} = \langle T\imath \circ T\nu U\phi_1, V\phi_1 \rangle_{L^2(P)} = \langle T\nu U, V \rangle_P, \tag{4.4}$$

which shows that $T\nu U$ is the projection of U onto $T_P N$ in the metric of (3.11). (This corresponds, in the ϕ_1-representation, to projection from $L^2(P)$ onto the subspace of functions with P-mean zero.) More generally, $T\nu$ effects 1-parallel transport of tangent vectors from $P \in M$ to $\nu(P) \in N$, followed by projection onto $T_{\nu(P)} N$.

The *Fisher-Rao metric* on TN is the restriction of the metric of (3.11) to TN:

$$\langle U, V \rangle_P = \langle U\phi_1, V\phi_{-1} \rangle_{L^2(\mu)} = \langle U\phi_1 - \mathbf{E}_\mu U\phi_1, V\phi_m \rangle_{L^2(\mu)} = \langle U\phi_e, V\phi_m \rangle_{L^2(\mu)}. \tag{4.5}$$

The α-covariant derivative on TN is the projection of that defined on TM in (3.9); for any $\mathbf{U}, \mathbf{V} \in \Gamma TN$,

$$\nabla_{\mathbf{U}}^\alpha \mathbf{V} = T\nu \circ \Phi_\alpha^{-1}(\phi_\alpha \circ \imath, \mathbf{U}\mathbf{V}(\phi_\alpha \circ \imath)). \tag{4.6}$$

Proposition 4.1. *For any $\alpha \in \mathbb{R}$ and any $\mathbf{U}, \mathbf{V} \in \Gamma TN$,*

$$\nabla_{\mathbf{U}}^\alpha \mathbf{V}\phi_e = \mathbf{U}\mathbf{V}\phi_e + \tfrac{1-\alpha}{2}\left[(\mathbf{U}\phi_e - \mathbf{E}_P \mathbf{U}\phi_e) \cdot (\mathbf{V}\phi_e - \mathbf{E}_P \mathbf{V}\phi_e) - \mathbf{E}_\mu(\mathbf{U}\phi_e - \mathbf{E}_P \mathbf{U}\phi_e) \cdot (\mathbf{V}\phi_e - \mathbf{E}_P \mathbf{V}\phi_e)\right]. \tag{4.7}$$

Proof. Let $\mathbf{W} \in \Gamma TM$ be defined by $\mathbf{W}_P \phi_\alpha = \mathbf{U}_{\nu(P)} \mathbf{V}(\phi_\alpha \circ \imath)$. According to (3.6) and (4.2), for any $P \in N$,

$$\begin{aligned}
\mathbf{W}_P \phi_1 &= (\phi_1 \circ \phi_\alpha^{-1})_{\phi_\alpha(P)}^{(1)} \mathbf{U}_P \left[(\phi_\alpha \circ \phi_1^{-1})_{\phi_1}^{(1)} (\phi_1 \circ \phi_e^{-1})_{\phi_e}^{(1)} \mathbf{v}_e\right] \\
&= \exp_{\mathbb{G}}\left(\tfrac{\alpha-1}{2}\phi_1\right) \cdot \mathbf{U}\left[\exp_{\mathbb{G}}\left(\tfrac{1-\alpha}{2}\phi_1\right) \cdot (\mathbf{v}_e - \mathbf{E}_{\mathrm{id}_N} \mathbf{v}_e)\right] \\
&= \tfrac{1-\alpha}{2}\mathbf{U}\phi_1 \cdot (\mathbf{v}_e - \mathbf{E}_P \mathbf{v}_e) + \mathbf{U}(\mathbf{v}_e - \mathbf{E}_{\mathrm{id}_N} \mathbf{v}_e) \\
&= \tfrac{1-\alpha}{2}(\mathbf{u}_e - \mathbf{E}_P \mathbf{u}_e) \cdot (\mathbf{v}_e - \mathbf{E}_P \mathbf{v}_e) + \mathbf{U}(\mathbf{v}_e - \mathbf{E}_{\mathrm{id}_N} \mathbf{v}_e),
\end{aligned}$$

where $\mathbf{u}_e := \mathbf{U}\phi_e$ and $\mathbf{v}_e := \mathbf{V}\phi_e$. Now $\nabla_{\mathbf{U}}^{\alpha}\mathbf{V} = T\nu\mathbf{W}$, and so $\nabla_{\mathbf{U}}^{\alpha}\mathbf{V}\phi_e = \mathbf{W}\phi_1 - \mathbf{E}_\mu\mathbf{W}\phi_1$, which completes the proof. $\square$

Remark 4.2. The *Amari-Chentsov tensor* on N is the symmetric covariant 3-tensor field τ, defined by

$$\tau_P(U, V, W) = \mathbf{E}_P(u - \mathbf{E}_P u) \cdot (v - \mathbf{E}_P v) \cdot (w - \mathbf{E}_P w), \tag{4.8}$$

where $U, V, W \in T_P N$, $u = U\phi_e$, $v = V\phi_e$ and $w = W\phi_e$. As in the finite-dimensional case,

$$\mathcal{D}_\alpha(\phi_e^{-1}|\phi_e^{-1})_{a,a}^{(2,1)}(u, v; w) = -\frac{1-\alpha}{2}\tau_P(U, V, W), \tag{4.9}$$

where $a = \phi_e(P)$, and this could be used to define the α-covariant derivative directly on N.

It follows from (4.7) that an α-geodesic of N is a smooth curve $\mathbf{P}$ satisfying the differential equation

$$\phi_e(\mathbf{P})'' = -\tfrac{1-\alpha}{2}\left[(\phi_e(\mathbf{P})' - \mathbf{E}_{\mathbf{P}}\phi_e(\mathbf{P})')^2 - \mathbf{E}_\mu(\phi_e(\mathbf{P})' - \mathbf{E}_{\mathbf{P}}\phi_e(\mathbf{P})')^2\right]. \tag{4.10}$$

The Fenchel-Legendre transform of Proposition 3.2 is preserved on N when $\alpha = \pm 1$; the role of the adjoint variables ϕ_1 and ϕ_{-1} is then played by ϕ_e and ϕ_m.

Setting $\alpha = 1$ in (4.7), we see that N is 1-flat and that ϕ_e is an affine chart for ∇^1. Furthermore, it is clear that N is also -1-flat and that ϕ_m is an affine chart for ∇^{-1}. N is thus *dually flat* ($\alpha = \pm 1$). Its -1-flatness arises from the trivial nature of its embedding in M when expressed in terms of the chart ϕ_{-1}; this is the natural embedding of a set of probability measures in a linear space of signed measures. Its 1-flatness is associated with its Lie group structure under the product $(PQ)_N := \nu(PQ)$.

The products on M and N have practical significance as "data fusion" operators in Bayesian estimation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, on which are defined random variables $X : \Omega \to B$ and $Y : \Omega \to \mathbb{Y}$, where $(\mathbb{Y}, \mathcal{Y}, \lambda)$ is a measure space. Let P_{XY} and P_X be the distributions of (X, Y) and X, respectively, and suppose that $P_X \in N$ and that there exists a measurable function $\Lambda : B \times \mathbb{Y} \to (0, \infty)$ with the following properties:

(i) $P_{XY}(A) = \int_A \Lambda \, d(P_X \otimes \lambda)$ for all $A \in \mathcal{X}_B \times \mathcal{Y}$;
(ii) $\Lambda(\cdot, y) \in \mathbb{G}^+$ for all $y \in \mathbb{Y}$.

For each y, $\Lambda(\cdot, y)$ is a *likelihood function* in the Bayesian problem of estimating X on the basis of the prior distribution P_X and the observation $Y = y$. Let $R : \mathbb{Y} \to M$ and $P_{X|Y} : \mathbb{Y} \to N$ be defined as follows:

$$R(y)(A) = \int_A \Lambda(x, y) \, \mu(\mathrm{d}x), \quad P_{X|Y}(y)(A) = \nu(P_X R(y)) = (P_X \nu(R(y)))_N; \tag{4.11}$$

then $P_{X|Y}(y)$ is a regular $(Y = y)$-conditional probability distribution for X. (A specific example is that in which $\mathbb{Y} = \mathbb{R}$, $Y = f(X) + \zeta$, $f \in \mathbb{G}$, and ζ has the standard Gaussian distribution and is independent of X. The likelihood function then takes the form $\Lambda(x, y) = n(0, 1)(y - f(x))$, where $n(0, 1)$ is the standard Gaussian density.)

The space $\mathbb{Y}$ is unimportant in estimation problems of this type in the sense that the observation $Y = y$ is represented by any of the surrogates $\Lambda(\cdot, y) \in \mathbb{G}^+$, $R(y) \in M$ or $\nu(R(y)) \in N$. Bayes' formula can thus be interpreted in terms of products of measures in the commutative Banach algebras M and N, making it the classical equivalent of the "collapse of state" of quantum probability, under observation. The latter involves the normalised product (iteration) of a quantum state (positive semi-definite operator) and a real-valued observable (Hermitian operator), both acting on an underlying Hilbert space [18].

Straightforward calculations show that, for any $\alpha \in \mathbb{R}$ and any $\mathbf{U}, \mathbf{V}, \mathbf{W} \in \Gamma T N$,

$$\nabla_{\mathbf{U}}^{\alpha}\nabla_{\mathbf{V}}^{\alpha}\mathbf{W} - \nabla_{\mathbf{V}}^{\alpha}\nabla_{\mathbf{U}}^{\alpha}\mathbf{W} - \nabla_{[\mathbf{U},\mathbf{V}]}^{\alpha}\mathbf{W} = \tfrac{1-\alpha^2}{4}\left(\langle\mathbf{V}, \mathbf{W}\rangle_P \mathbf{U} - \langle\mathbf{U}, \mathbf{W}\rangle_P \mathbf{V}\right).$$

The *Riemann curvature tensor* for the α-covariant derivative, R^α, is thus the covariant 4-tensor field defined by

$$R_P^\alpha(U,V,W,X) = \tfrac{1-\alpha^2}{4}\left(\langle V,W\rangle_P \langle U,X\rangle_P - \langle U,W\rangle_P \langle V,X\rangle_P\right), \tag{4.12}$$

where $U,V,W,X \in T_P N$. This is equal for the dual covariant derivatives $\nabla^{\pm\alpha}$, and is zero if and only if $\alpha = \pm 1$.

Since the statistical manifold N is covered by a single chart, it is flat with respect to infinitely many covariant derivatives. If $\gamma : \mathbb{G}_0 \to \mathbb{Y}$ is a diffeomorphism into a Banach space $\mathbb{Y}$, then $\gamma \circ \phi_e : N \to \mathbb{Y}$ is another global chart, which induces its own parallel transport. An interesting example is the Tsallis *q-deformed logarithmic chart s_q*, as used in [16]. In the present context, this has range space $\mathbb{G}_0$ and is defined for any $q \neq 1$ by

$$s_q(P) = (1-q)^{-1}\phi_m(\nu(P^{1-q})). \tag{4.13}$$

The associated covariant derivative $\tilde\nabla^q$ is that for which $\tilde\nabla_{\mathbf{U}}^q \mathbf{V}s_q = \mathbf{U}\mathbf{V}s_q$. Expressed in terms of the chart ϕ_e:

$$\tilde\nabla_{\mathbf{U}}^q \mathbf{V}\phi_e = \mathbf{U}\mathbf{V}\phi_e + (1-q)\left[(\mathbf{u}_e - \mathbf{E}_Q\mathbf{u}_e)(\mathbf{v}_e - \mathbf{E}_Q\mathbf{v}_e) - \mathbf{E}_\mu(\mathbf{u}_e - \mathbf{E}_Q\mathbf{u}_e)(\mathbf{v}_e - \mathbf{E}_Q\mathbf{v}_e)\right], \tag{4.14}$$

where $\mathbf{E}_Q$ is expectation with respect to the probability measure $Q := \nu(P^{1-q})$. Like the manifold constructed in [16], N is $\tilde\nabla^q$-flat; however, unless $q = 0$ (in which case $\tilde\nabla^q = \nabla^{-1}$), $\tilde\nabla^q$ differs from the α-covariant derivatives. If N' is a second copy of N, and $F_q : N \to N'$ is defined by $F_q(P) = \nu(P^{1-q})$, then the $\tilde\nabla^q$-parallel transport on N is the pullback of the ∇^{-1}-parallel transport on N'.

5. Manifolds of smooth densities

In this section we consider the *sequences* of manifolds $(M^k, k \in \mathbb{N}_0)$ and $(N^k, k \in \mathbb{N}_0)$, as developed in Sections 3 and 4, making explicit their dependence on the number of derivatives in the definition of $\mathbb{G}$ $(= \mathbb{G}^k)$. By developing projective limits of these sequences, we define Fréchet manifolds of measures having smooth densities with respect to μ. The manifold of finite measures in this context, its tangent bundle, and its model space, are as follows:

$$\bar{M} := \bigcap_{k\in\mathbb{N}_0} M^k, \quad T\bar{M} := \bigcap_{k\in\mathbb{N}_0} TM^k \quad \text{and} \quad \bar{\mathbb{G}} := \bigcap_{k\in\mathbb{N}_0} \mathbb{G}^k. \tag{5.1}$$

Let $\rho^k : \bar{\mathbb{G}} \to \mathbb{G}^k$ be the inclusion map. $\bar{\mathbb{G}}$ is a Fréchet space, whose topology is generated by the sequence of norms $(\|\rho^k\|_{\mathbb{G}^k}, k \in \mathbb{N}_0)$; $\bar{M}$ is a Fréchet manifold of finite measures on $(B, \mathcal{X}_B)$, whose densities with respect to μ are smooth bounded functions, having bounded derivatives of all orders. Its tangent bundle is trivialised by the chart $\bar\Phi_1 : T\bar{M} \to \bar{\mathbb{G}} \times \bar{\mathbb{G}}$, which is defined to be the restriction of Φ_1^k to $T\bar{M}$. A tangent vector at $\bar{P} \in \bar{M}$ is a signed measure on $(B, \mathcal{X}_B)$, whose density with respect to $\bar{P}$ is of the form $d\bar{U}/d\bar{P} = u$ for some $u \in \bar{\mathbb{G}}$.

A map $f : \bar{\mathbb{G}} \to \bar{\mathbb{F}}$, taking values in another Fréchet space $\bar{\mathbb{F}}$, is said to be *Leslie differentiable* [14], with derivative $df : \bar{\mathbb{G}} \to L(\bar{\mathbb{G}}; \bar{\mathbb{F}})$, if, for any $a \in \bar{\mathbb{G}}$, the map $R_a : \mathbb{R} \times \bar{\mathbb{G}} \to \bar{\mathbb{F}}$ defined by

$$R_a(t,u) := \begin{cases} t^{-1}(f(a+tu) - f(a)) - df(a)u & \text{if } t \neq 0 \\ 0 & \text{if } t = 0, \end{cases} \tag{5.2}$$

is continuous at $(0,u)$ for every $u \in \bar{\mathbb{G}}$. The study of the Leslie differentiability properties of a map between Fréchet spaces (including the regularity of its derivatives, considered as maps into spaces of continuous linear maps) becomes substantially easier if the map in question is the projective limit of a system of maps between Banach spaces [10].

For any $0 \leq j \leq k < \infty$, let $\rho^{kj} : \mathbb{G}^k \to \mathbb{G}^j$ be the (continuous linear) inclusion map. The system $(\mathbb{G}^k, \rho^{kj}, 0 \leq j \leq k < \infty)$ is a *projective system* with *factor spaces* $\mathbb{G}^k$ and *connecting morphisms* ρ^{kj}. The *projective limit* of

this system is the following subset of the cartesian product $\Pi := \prod_{k=0}^{\infty} \mathbb{G}^k$:

$$\varprojlim \mathbb{G}^k := \left\{ (a^0, a^1, \ldots) \in \Pi : \rho^{kj} a^k = a^j \text{ for all } 0 \leq j \leq k < \infty \right\}. \qquad (5.3)$$

In this particular example, the map $\varprojlim \mathbb{G}^k \ni (\rho^0 \bar{a}, \rho^1 \bar{a}, \ldots) \mapsto \bar{a} \in \bar{\mathbb{G}}$ is a toplinear isomorphism, and so we can identify $\varprojlim \mathbb{G}^k$ with $\bar{\mathbb{G}}$. The inclusion map $\rho^k : \bar{\mathbb{G}} \to \mathbb{G}^k$ then plays the role of the *canonical projection* [10].

Suppose that $(\mathbb{F}^k, \sigma^{kj}, 0 \leq j \leq k < \infty)$ is another projective system of Banach spaces with projective limit $\bar{\mathbb{F}}$. The sequence $(f^k : \mathbb{G}^k \to \mathbb{F}^k, k \in \mathbb{N}_0)$ is a *projective system of maps* if

$$\sigma^{kj} f^k = f^j \rho^{kj} \quad \text{for all } 0 \leq j \leq k < \infty. \qquad (5.4)$$

The projective limit of this system is $\bar{f} : \bar{\mathbb{G}} \to \bar{\mathbb{F}}$, defined by $\bar{f}(\bar{a}) = (f^0(\bar{a}), f^1(\bar{a}), \ldots)$. If each f^k is (Fréchet) differentiable then $\bar{f}$ is Leslie differentiable, and its derivative can be associated with a projective limit of those of f^k (see Prop. 2.3.11 in [10]). The appropriate projective system of derivatives is $(\Delta f^k : \mathbb{G}^k \to H^k(\mathbb{G}; \mathbb{F}), \ k \in \mathbb{N}_0)$, where

$$\Delta f^k := \left(f^{0,(1)}_{\rho^{k0}}, f^{1,(1)}_{\rho^{k1}}, \ldots, f^{k,(1)} \right), \qquad (5.5)$$

and

$$H^k(\mathbb{G}; \mathbb{F}) = \left\{ (\lambda^0, \ldots, \lambda^k) \in \prod_{i=0}^{k} L(\mathbb{G}^i; \mathbb{F}^i) : \sigma^{ji} \lambda^j = \lambda^i \rho^{ji}, \ i \leq j \right\}. \qquad (5.6)$$

The factor spaces $H^k(\mathbb{G}; \mathbb{F})$ are connected by the morphisms $h^{kj} : H^k(\mathbb{G}; \mathbb{F}) \to H^j(\mathbb{G}; \mathbb{F})$,

$$h^{kj}(\lambda^0, \ldots, \lambda^k) = (\lambda^0, \ldots, \lambda^j), \quad j \leq k,$$

and so constitute a projective system of Banach spaces. The associated projective limit is toplinear isomorphic with $\bar{H}(\mathbb{G}; \mathbb{F})$ (defined by the obvious variant of (5.6)), and the map $\epsilon : \bar{H}(\mathbb{G}; \mathbb{F}) \to L(\bar{\mathbb{G}}; \bar{\mathbb{F}})$, defined by $\epsilon(\lambda^0, \lambda^1, \ldots) = \varprojlim \lambda^k = (\lambda^0 \rho^0, \lambda^1 \rho^1, \ldots)$, is continuous linear, with respect to the toplogy of uniform convergence on bounded sets (see Thm. 2.3.10 in [10]). That $\bar{H}(\mathbb{G}; \mathbb{F})$ is a projective limit of Banach spaces is central to the regularity of the Leslie derivative $d\bar{f}$. If each f^k is smooth then $d\bar{f} : \bar{\mathbb{G}} \to L(\bar{\mathbb{G}}; \bar{\mathbb{F}})$ is Leslie smooth (see Prop. 2.3.11, 2.3.12 in [10]).

Applying these ideas to the transition maps $\Phi_\alpha^k \circ (\Phi_1^k)^{-1}$ and their inverses, we see that the projective limit $\bar{\Phi}_\alpha \circ \bar{\Phi}_1^{-1}$ is Leslie diffeomorphic, and all its derivatives (together with those of its inverse) are smooth maps from open subsets of $\bar{\mathbb{G}}$ to appropriate spaces of continuous linear maps. $(\bar{\Phi}_\alpha, \alpha \in \mathbb{R})$ is a Leslie smooth atlas for $T\bar{M}$. For any Leslie differentiable map $f : \bar{\mathbb{G}} \to \bar{\mathbb{F}}$ and any $\bar{U} \in T_{\bar{P}} \bar{M}$

$$\bar{U} f = d(\bar{f} \circ \bar{\phi}_\alpha^{-1}) \bar{U} \bar{\phi}_\alpha \quad \text{for any } \alpha \in \mathbb{R}.$$

We define a special class of smooth vector fields of $\bar{M}$ – those whose $\bar{\Phi}_\alpha$-representations are projective limits of smooth maps between the Banach spaces $\mathbb{G}^k$. Let S be the following set of sequences:

$$S = \left\{ (\mathbf{n}_k \in \mathbb{N}_0, k \in \mathbb{N}_0) : \mathbf{n}_k \leq \mathbf{n}_{k+1}, \ \sup \mathbf{n}_k = +\infty \right\}, \qquad (5.7)$$

and note that, for any $\mathbf{n} \in S$, $(\mathbb{G}^{\mathbf{n}_k}, \rho^{\mathbf{n}_k \mathbf{n}_j}, 0 \leq j \leq k < \infty)$ is a projective system of Banach spaces with projective limit $\bar{\mathbb{G}}$. For some $\mathbf{n} \in S$, let $(\mathbf{u}^k : \mathbb{G}^k \to \mathbb{G}^{\mathbf{n}_k}, \ k \in \mathbb{N}_0)$ be a projective system of smooth maps, with projective limit $\bar{\mathbf{u}} : \bar{\mathbb{G}} \to \bar{\mathbb{G}}$. We regard $\bar{\mathbf{u}} \circ \bar{\phi}_1$ as being the $\bar{\Phi}_1$-representation of a smooth vector field $\bar{\mathbf{U}}$ $(\bar{\mathbf{U}} \bar{\phi}_1 := \bar{\mathbf{u}} \circ \bar{\phi}_1)$. We denote the set of all such *projective-limit smooth vector fields* $\Gamma_{\mathrm{pl}} T\bar{M}$. This has a linear

structure, in which the sum of $(\mathbf{u}^k : \mathbb{G}^k \to \mathbb{G}^{\mathbf{n}_k})$ and $(\mathbf{v}^k : \mathbb{G}^k \to \mathbb{G}^{\mathbf{m}_k})$, for $\mathbf{m}, \mathbf{n} \in S$, is the projective system $(\mathbf{w}^k := \rho^{\mathbf{n}_k \mathbf{l}_k} \mathbf{u}^k + \rho^{\mathbf{m}_k \mathbf{l}_k} \mathbf{v}^k : \mathbb{G}^k \to \mathbb{G}^{\mathbf{l}_k})$, where $\mathbf{l}_k := \min\{\mathbf{m}_k, \mathbf{n}_k\}$.

Remark 5.1. $\Gamma_{\text{pl}} T\bar{M}$ is strictly smaller than $\Gamma T\bar{M}$ – it does not contain the vector field with $\bar{\Phi}_1$-representation $\bar{\mathbf{u}}(\bar{a}) = \bar{a} r(\bar{a}, 0)$, for example, where r is the usual metric on $\bar{\mathbb{G}}$. However, it does contain many useful vector fields occurring in the theory of partial differential equations. For example, if $\mathbb{X} = \mathbb{R}^d$ then the second-order differential operator $\partial^2 / \partial x_i \partial x_j$ lifts to a vector field in $\Gamma_{\text{pl}} T\bar{M}$. (*cf.* (1.5).)

Proposition 5.2. *Let* $\bar{\mathbf{u}} : \bar{\mathbb{G}} \to \bar{\mathbb{G}}$ *be as defined above, and let* $(f^k : \mathbb{G}^k \to \mathbb{F}^k)$ *be a projective system of* smooth *maps, as described in (5.4). Then the sequence of maps*

$$\left(f^{l,(1)}_{\rho^{kl}} \rho^{\mathbf{n}_k l} \mathbf{u}^k : \mathbb{G}^k \to \mathbb{F}^l, \ l = \min\{k, \mathbf{n}_k\}, \ k \in \mathbb{N}_0 \right) \tag{5.8}$$

is a projective system, with projective limit $\mathrm{d}\bar{f}\bar{\mathbf{u}}$, *and* $\bar{\mathbf{U}}(\bar{f} \circ \bar{\phi}_1) = \mathrm{d}\bar{f}\bar{\mathbf{u}} \circ \bar{\phi}_1$.

Proof. For any $j \leq k$, let $l := \min\{k, \mathbf{n}_k\}$ and $m := \min\{j, \mathbf{n}_j\}$. Differentiating the projective relation $\sigma^{lm} f^l = f^m \rho^{lm}$, we obtain $\sigma^{lm} f^{l,(1)} = f^{m,(1)}_{\rho^{lm}} \rho^{lm}$. Restricting the base-point from $\mathbb{G}^l$ to $\mathbb{G}^k$, and applying the resulting linear map to $\rho^{\mathbf{n}_k l} \mathbf{u}^k$, we obtain

$$\sigma^{lm} f^{l,(1)}_{\rho^{kl}} \rho^{\mathbf{n}_k l} \mathbf{u}^k = f^{m,(1)}_{\rho^{km}} \rho^{\mathbf{n}_k m} \mathbf{u}^k = \left(f^{m,(1)}_{\rho^{jm}} \rho^{\mathbf{n}_j m} \mathbf{u}^j \right) \rho^{kj},$$

which establishes the projective property. The projective limit is

$$\left(f^{\mathbf{l}_0,(1)}_{\rho^{\mathbf{l}_0}} \rho^{\mathbf{l}_0} \bar{\mathbf{u}}, f^{\mathbf{l}_1,(1)}_{\rho^{\mathbf{l}_1}} \rho^{\mathbf{l}_1} \bar{\mathbf{u}}, \dots \right) \equiv \mathrm{d}\bar{f}\bar{\mathbf{u}},$$

where $\mathbf{l}_k := \min\{k, \mathbf{n}_k\}$. Let $\mathbf{P} \in \bar{\mathbf{U}}(\bar{P})$; then

$$\bar{\mathbf{U}}(\bar{P})(\bar{f} \circ \bar{\phi}_1) = (\bar{f} \circ \bar{\phi}_1(\mathbf{P}))'(0) = \mathrm{d}\bar{f}\bar{\mathbf{U}}(\bar{P})\bar{\phi}_1 = \mathrm{d}\bar{f}\bar{\mathbf{u}} \circ \bar{\phi}_1(\bar{P}),$$

which completes the proof.   □

Suppose that $\bar{\mathbf{V}} \in \Gamma_{\text{pl}} T\bar{M}$ is defined by the projective system of smooth maps $(\mathbf{v}^k : \mathbb{G}^k \to \mathbb{G}^{\mathbf{m}_k}, \ k \in \mathbb{N}_0)$ for some $\mathbf{m} \in S$. By applying Proposition 5.2 to the projective system $(\phi_\alpha^{\mathbf{m}_k} \circ (\phi_1^{\mathbf{m}_k})^{-1} \circ \mathbf{v}^k : \mathbb{G}^k \to \mathbb{G}^{\mathbf{m}_k} (=: \mathbb{F}^k))$, we can define the α-covariant derivative on $\bar{M}$: $\nabla^\alpha_{\bar{\mathbf{U}}} \bar{\mathbf{V}} \bar{\phi}_1 = \bar{\mathbf{w}} \circ \bar{\phi}_1$, where

$$\bar{\mathbf{w}} = d(\bar{\phi}_1 \circ \bar{\phi}_\alpha^{-1}) d(d(\bar{\phi}_\alpha \circ \bar{\phi}_1^{-1})\bar{\mathbf{v}})\bar{\mathbf{u}} = \mathrm{d}\bar{\mathbf{v}}\bar{\mathbf{u}} + \tfrac{1-\alpha}{2} \bar{\mathbf{v}} \cdot \bar{\mathbf{u}}. \tag{5.9}$$

The $\bar{\Phi}_1$-representation $\bar{\mathbf{w}}$ is the projective limit of the system $(\mathbf{w}^k : \mathbb{G}^k \to \mathbb{G}^{\mathbf{i}_k}, \ k \in \mathbb{N}_0)$, where $(\mathbf{i}_k = \min\{\mathbf{m}_k, \mathbf{n}_k, \mathbf{m}_{\mathbf{n}_k}\}, \ k \in \mathbb{N}_0) \in S$, and so $\nabla^\alpha_{\bar{\mathbf{U}}} \bar{\mathbf{V}} \in \Gamma_{\text{pl}} T\bar{M}$. The remaining constructions in Sections 3 and 4 carry over to $\bar{M}$ without difficulty. Key points to note are as follows.

- The smoothness of the α-divergences on $\bar{M}$ follows from their smoothness on M^k, and that of the inclusion map $\imath^k : \bar{M} \to M^k$. The metric and covariant derivatives could be derived directly from $\mathcal{D}_\alpha$ as in Sections 3 and 4.
- The statistical manifold $\bar{N}$ is defined in the obvious way. It is a Leslie C^∞-embedded submanifold of $\bar{M}$ since its image under $\bar{\phi}_{-1}$ is the subspace of $\bar{\mathbb{G}}$ comprising those members with zero μ-mean.
- An α-geodesic of $\bar{N}$ is a smooth curve $\mathbf{P}$ whose projection $\imath^k \mathbf{P}$ satisfies (4.10) for all k. (α-geodesics of $\bar{M}$, and ± 1-geodesics of $\bar{N}$ are, of course, straight lines in appropriate charts.)

6. CONCLUDING REMARKS

In this paper we have constructed a family of non-parametric statistical manifolds, N, that support the full geometry of the Fisher-Rao metric and Amari α-covariant derivatives for all $\alpha \in \mathbb{R}$. Manifolds in the family admit global mixture and exponential charts, ϕ_m and ϕ_e, which are of importance in applications. The α-covariant derivatives were computed in the chart ϕ_e, and their curvature on N was found: N is α-flat if and only if $\alpha = \pm 1$; otherwise the curvature tensor changes sign as $|\alpha|$ passes through 1. The α-divergences are of class C^∞ on N. As in the parametric case, their derivatives provide an alternative way of defining covariant derivatives.

The statistical manifolds were constructed *extrinsically*, *via* smooth embeddings in particular manifolds of finite measures, M. The latter are covered by every chart in a smooth one-parameter atlas ($\phi_\alpha, \alpha \in \mathbb{R}$), each chart inducing its own parallel transport on the tangent bundle. Since $M = \phi_1^{-1}(\mathbb{G})$ and $N = \phi_e^{-1}(\mathbb{G}_0)$, the manifolds M and N are, in one sense, no more than linear spaces themselves. The statistical interest comes from the interplay between the different affine representations provided by the charts ϕ_α, in particular ϕ_1 and ϕ_{-1}. Manifold theory enters the picture with the introduction of the base-point-dependent Fisher-Rao metric on the tangent bundle.

The manifolds are applicable to problems in Bayesian estimation, in which the α-divergences are important measures of approximation error. In particular, the process of posterior distributions of a nonlinear filter can be expressed as a solution of a suitable differential equation on N. In this, and other potential applications to Physics (*eg.* the Fokker-Planck or the Boltzmann equations), it is important for the members of N to have differentiable densities, and this is built in to its definition. A Fréchet manifold of probability measures having *smooth* densities was defined in Section 5, *via* projective limits. Under suitable technical conditions, the coefficients of certain partial differential equations, including those of nonlinear filtering, can be interpreted as vector fields of this manifold.

The extra regularity of N, over other non-parametric manifolds in the literature, is gained at the cost of inclusiveness: each probability density in N has a strictly positive infimum. Although many idealised models of physical systems do not exhibit this property, it is often possible to substitute a model that does. In the context of nonlinear filtering of diffusion processes, for example, we can constrain the diffusion to stay on a bounded domain by introducing suitable boundary conditions. It is open to question whether this results in any less accurate a model of reality than the idealised model. This is a question for future research. Another potential avenue is to use N as a starting point in the construction of manifolds that place less stringent regularity constraints on their members.

REFERENCES

[1] S.-I. Amari and H. Nagaoka, Methods of Information Geometry. Vol. 191 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence (2000).

[2] N. Ay, J. Jost, H.V. Lê and L. Schwachhofer, Information geometry and sufficient statistics. *Probab. Theory Relat. Fields* **162** (2015) 327–364.

[3] N. Ay, J. Jost, H.V. Lê and L. Schwachhofer, Paramerized Measure Models. *Bernoulli* **24** (2018) 1692-1725.

[4] O.E. Barndorff-Nielsen, Information and Exponential Families in Statistical Theory. Wiley (1978).

[5] M. Bauer, M. Bruveris and P.W. Michor, Uniqueness of the Fisher-Rao metric on the space of smooth densities. *Bull. Lond. Math. Soc.* **48** (2016) 499–506.

[6] D. Brigo and G. Pistone, Dimensionality reduction for measure valued evolution equations in statistical manifolds, in Computational Information Geometry for Image and Signal Processing, edited by F. Nielsen, F. Critchley and C.T.J. Dodson. Springer (2017) 217–265.

[7] M. Bruveris and P.W. Michor, Geometry of the Fisher-Rao Metric on the Space of Smooth Densities on a Compact Manifold. Preprint arXiv:1607.04550 (2016).

[8] A. Cena and G. Pistone, Exponential statistical manifold. *Ann. Inst. Stat. Math.* **59** (2007) 27–56.

[9] N.N. Chentsov, Statistical Decision Rules and Optimal Inference. Vol. 53 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence (1982).

[10] C.T.J. Dodson, G. Galanis and E. Vassiliou, Geometry in a Fréchet Context: A Projective Limit Approach. Vol. 428 of *London Mathematical Society Lecture Note Series*. Cambridge University Press (2016).

[11] K. Fukumizu, Exponential manifold by reproducing kernel Hilbert spaces, in Algebraic and Geometric Methods in Statistics, edited by P. Gibilisco, E. Riccomagno, M.P. Rogantin and H. Winn. Cambridge University Press (2009) 291–306.

[12] P. Gibilisco and G. Pistone, Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **1** (1998) 325–347.

[13] S.L. Lauritzen, Statistical Manifolds. Vol. 10 of *IMS Lecture Notes Series*. Institute of Mathematical Statistics (1987).

[14] J.A. Leslie, On a differential structure for the group of diffeomorphisms. *Topology* **46** (1967) 263–271.

[15] R.S. Liptser and A.N. Shiryayev, Statistics of Random Processes I – General Theory. Springer (2001).

[16] G. Loaiza and H.R. Quiceno, A *q*-exponential statistical Banach manifold. *J. Math. Ann. Appl.* **398** (2013) 466–476.

[17] B. Lods and G. Pistone, Information geometry formalism for the spatially homogeneous Boltzmann equation. *Entropy* **17** (2015) 4323–4363.

[18] P.A. Meyer, Quantum Probability for Probabilists. Vol. 1538 of *Lecture Notes in Mathematics*. Springer (1995).

[19] M.K. Murray and J.W. Rice, Differential Geometry and Statistics. Vol. 48 of *Monographs in Statistics and Applied Probability*. Chapman Hall (1993).

[20] J. Naudts, Generalised Thermostatistics. Springer, London (2011).

[21] N.J. Newton, An infinite-dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.* **263** (2012) 1661–1681.

[22] N.J. Newton, Information geometric nonlinear filtering. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **18** (2015) 1550014.

[23] N.J. Newton, Infinite-dimensional statistical manifolds based on a balanced chart. *Bernoulli* **22** (2016) 711–731.

[24] F. Nielsen and F. Barbaresco, Proceedings of GSI 2013 Conference. Vol. 8085 of *Lecture Notes in Computer Science*. Springer, Berlin (2013).

[25] F. Nielsen and F. Barbaresco, Proceedings of GSI 2015 Conference. Vol. 9389 of *Lecture Notes in Computer Science*. Springer, Berlin (2015).

[26] G. Pistone and M.P. Rogantin, The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli* **5** (1999) 721–760.

[27] G. Pistone and C. Sempi, An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Stat.* **23** (1995) 1543–1561.

[28] C.R. Rao, Information and accuracy obtainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37** (1945) 81–91.